

Future Research in the Social, Behavioural and Economic Sciences: addressing data needs

Peter Elias
Strategic Advisor
UK Economic and Social Research Council

October 2010

Abstract

This paper addresses the link between future research challenges and the data infrastructure required to address these challenges. It draws on a wide consultation held with the UK research community in the social, behavioural and economic sciences undertaken in 2009 to inform the strategic development of such resources. It stresses the importance of renewed efforts to improve access to existing data on a global basis and for further work to explore the potential research value of new forms of data for research purposes.

Creative Commons licence details

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Introduction

Research lies at the heart of innovation and knowledge creation. Over the past 20 years, with the growing pace of innovation and the associated transfer of knowledge from research to practical application, research activity in all of the sciences has both broadened and deepened. National and international collaborations between research teams, multidisciplinary approaches to research issues and the development of large-scale research programmes have become common place. But it is not just the desire to innovate or to pursue ‘knowledge for knowledge’s sake’ that has stimulated research activity. There is now a growing recognition of the challenges facing human populations which require more intensive research efforts and wider research collaboration.

Across the social, economic and behavioural sciences, the issues that demand focussed research efforts have become ever more apparent, ranging from the need to understand instability in economic and financial markets, the links between socio-economic behaviour and environmental changes, population movements in response to poverty, war, disease and natural disasters and the complex relationships between mental and physical health, family development, education and lifestyles.

Research in all of these areas requires data – records and measurements taken from the physical and the human environment. Research data are more than just numbers. They may be quantitative or qualitative, deriving from specialist surveys, experimental techniques, the processing of transactions or from sensing devices. They may consist of numerical measurements or written text, visual images and audio recordings. Technological advances now provide the opportunity to capture and preserve digital information on a scale which would have been unimaginable just two decades ago. The vast increase in information arising from the digital revolution has the potential to improve further and accelerate research efforts, provided that the requisite data resources needed for scientific research can be collected, marshalled and preserved in ways that facilitate high quality research.

UK strategic approach to the identification and development of data resources for social and economic research

Research data resources, whether created specifically for research purposes or generated as the by-product of other activities, are expensive to develop and maintain and often require extensive collaboration between different agencies. The *UK Strategy for Data Resources for Social and Economic Research* (the ‘National Data Strategy’) addresses these problems by seeking to identify, prioritise and assist with the development of research data in the light of pressing research needs. The strategy helps shape the way in which UK-based organisations with interests in social science research data (data producers, research funding bodies, data archives, *etc.*) work together and with the research community to maximise the research potential of existing data and to create new resources, developing better access to existing data and facilitating a broad research agenda.

The challenges driving data needs

Data needs are driven by research and policy interests. Whether for description or more detailed analysis, measurements taken either from the world around us or about the human condition form key parts of the processes of scientific enquiry and policy formulation. The task of determining and prioritising data needs must be linked to the potential research and policy agenda.

This ‘research and policy driven’ approach guided the development of the first UK National Data Strategy covering the period 2006 to 2009. Four key research challenges were identified, each of which gave rise to data requirements of various types from a range of sources. These related to: the ageing population; migration; globalisation; and child development. A number of important data resources that could inform research on these issues were recognised, including: large-scale census and survey data; longitudinal data on individuals, families and households; administrative data relating to health, social security, employment and education; and data on organisations. For each data type, the importance of both geographical location of activity and the timeliness of the data were highlighted.

Following an on-line user consultation and the workshops with key data stakeholders (users and producers of data resources) designed to inform the first revision of the strategy, concerns were expressed about the need to adopt a broader and more flexible approach to the identification of future research challenges. In particular, it was argued that the first UK National Data Strategy did not reflect the growing importance which attaches to the internationalisation of all aspects of the research agenda. Increased concerns about threats to global security, large-scale migration, the global impact of pressures in financial markets and the fluctuating demands for energy, basic materials such as coal, steel and copper from emerging economies, and global health issues all seem likely to form an important part of the international research agenda over the next few years. Similarly, global efforts to achieve major reductions in anthropogenic CO₂ raise issues about trade and competitiveness that will stimulate the need for more and better data on labour and raw material costs, productivity and trade flows. More generally, many stakeholders argued that social scientists must increase their engagement with issues related to environmental changes, whether in response to climate change, pollution, and resource management or in terms of the need to promote and sustain changes in behaviour.

Other specific concerns include the need for improved information which assists with the production of population estimates at the detailed spatial level, for better statistics on crime and crime-related behaviour and for significant improvements in access to high quality data to facilitate research on all aspects of health and well-being.

Addressing data needs associated with these challenges will require cooperation and collaboration between various departments and agencies, with support from research funders, to achieve linkage between data sources at the individual, organisational and geographical levels. In turn, this requires that the issues surrounding the ethics of linkage

and the security, preservation and sustainability of the linked data so created are addressed and resolved in ways which maintain public support for these efforts and enhance the research value of the data resources so created.

The digital revolution – proving research value and setting priorities

Technological innovation, particularly the development of digital data capture techniques, the vast increase in data storage capacity at very low cost and mobile communications, has brought about what is often termed the ‘data deluge’ – the voluminous amounts of information which record such things as the everyday movements of people, the transactions they conduct and the state of the environment they inhabit. The information so generated is not primarily designed for research but addresses the needs of organisations to record, monitor and track activities in the interests of commercial or administrative efficiency and reliability.

Some indication of the research value of digital records can be gained by grouping into four broad categories the various types of digital data which have research potential:

- *Administrative data*- a term usually used for data which arise through the administration of a process within the public sector (*e.g.* a claim for social benefits, patient health records, student educational progress records, tax records).
- *Commercial data* – are generated by organisations usually operating on a ‘for-profit’ basis and are used primarily to improve the efficiency and/or profitability of an organisation. (*e.g.* loyalty card records, utility usage, bank and financial records).
- *Tracking data* – may be a hybrid of the two above types (*e.g.* mobile telephone logs) or may be generated by remote sensing devices (*e.g.* CCTV footage, vehicle movement sensors, web page hits).
- *User-generated data* – generally refers to data requested from users of a specific technology, allowing databases of information to be built up quickly and reliably at very low cost. A good example is the generation of maps for Haiti immediately after the earthquake, initially from satellite images, but placed in a format which allowed users to ‘tag’ detail and add content to be shared by all.

Data sharing has become common place within the physical sciences. However, social science data, like medical data, present specific issues that the physical sciences do not usually need to confront. What differentiates such data are their personal nature – the fact that measurements relate to people and/or organisations. This, in turn, creates problems relating to the need to protect identities, the difficulty of gaining access to subjects, the need to respect the wishes and expectations of those subjects, the political sensitivity of the information collected and the associated higher research costs involved in resolving these problems whilst facilitating high quality research. There is, therefore, a tension between the difficulties of using appropriate methods to capture, process and store data about people and organisations, the problems of prioritising data needs given the sheer volume of information now potentially available for research, the need to preserve security and protect the identities of those persons and organisations to whom

the data relate and the need to ensure that commercial sensitivities and the wishes of individuals are respected.

Governments, research funding agencies and data regulatory authorities in many countries have been grappling with issues that relate to the research potential arising from the digital information revolution. The UK National Data Strategy recognises the latent research potential in these types of data. Steps will be taken to establish and promote mechanisms to explore their research value whilst overcoming the practical issues associated with reuse.

A strategy for international data requirements

One of the strongest themes arising from the consultations with producers and users of data relates to the growing importance which attaches to the internationalisation of the research agenda. There is now a widespread awareness of the need for social scientists, often in collaboration with researchers from other research disciplines, to take a more global approach to many major issues. These include a better understanding of the interdependencies between financial markets and world trade, research on the impacts of global cooperation to reduce CO₂ emissions, the socio-economic consequences of climate change, patterns of migration and their impact on communities, the spread of infectious diseases, and the global security challenges arising from poverty, war and political instability. Research on these issues requires collaboration across national boundaries involving knowledge exchange, transfer of research expertise and data sharing.

Significant work has already been undertaken to identify and provide access to a wide range of data resources which can inform a comparative research agenda. While these efforts must continue to be supported, there is now growing concern about the problems that researchers face when seeking to collaborate across national boundaries on issues of global concern. The problems relate to the difficulties of discovering relevant data, of gaining research access to detailed and up-to-date data, of ascertaining the quality of data and of preserving data for reuse. Language barriers often exacerbate these problems despite the growing availability of translation tools.

Steps must be taken to progress international collaboration at the highest levels over issues relating to the sharing of data resources, the resolution of legal problems relating to access to data and the promotion of research networks with the skills and knowledge to make use of the range of data resources potentially available world-wide. Given the need for international cooperation in these areas, these initiatives will require substantial efforts from a number of countries if they are to bear fruit.

Conclusion

Much work has been identified as of high priority if the UK, together with its research funding counterparts in other countries, is to address data needs associated with the major research challenges in the social, economic and behavioural sciences. Two common themes have emerged from the recent consultation. First, efforts must be

made to improve global access to data that already exist, but which cannot be accessed for research purposes due to legal, technical, ethical, cultural or political considerations. Sharing data for publically funded research purposes should be the norm rather than the exception, which is too often the case at present. Secondly, new forms of data are arising which may have vast research potential. Research and data scientists need to work together at the international level to explore and exploit this potential.